DOCUMENT RESUME

ED 334 339 UD 028 189

AUTHOR Darling-Hammond, Linda; Ascher, Carol
TITLE Creating Accountability in Big City School Systems. Urban Diversity Series No. 102.
INSTITUTION Columbia Univ., New York, NY. Teachers Coll. National Center for Restructuring Education, Schools and Teaching.; ERIC Clearinghouse on Urban Education, New York, N.Y.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Mar 91
CONTRACT RI88062013
NOTE 48p.; On title page, cited as "Creating Accountability in Big City Schools."
AVAILABLE FROM ERIC Clearinghouse on Urban Education, Teachers College, Box 40, Columbia University, New York, NY 10027.
PUB TYPE Information Analyses - ERIC Clearinghouse Products (071)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; *Educational Assessment; *Educational Improvement; Elementary Secondary Education; Evaluation Criteria; Outcomes of Education; School Effectiveness; School Restructuring; *Systems Development; *Urban Schools
IDENTIFIERS *Educational Indicators

ABSTRACT

Accountability has always been a basic concept in public education, although ideas about how to accomplish it have changed over the years. Problems in urban schools have given rise to the hope that carefully created systems of accountability might spur school improvement and school restructuring. Devising a system of genuine accountability in a large urban school is a complex task, involving careful sorting of responsibilities and a thoughtful set of measures for assessing school effectiveness and student progress. The following types of mechanisms operate simultaneously within a system of accountability: (1) political accountability; (2) legal accountability; (3) bureaucratic accountability; (4) professional accountability; and (5) market accountability. Bureaucratic accountability, professional accountability, and market accountability are all currently proposed as strategies for school improvement. Accountability systems need multiple statistical indicators to stimulate and measure school improvement; however, there is little agreement about what those indicators should be, or who should be responsible for determining them. Indicators must go beyond rudimentary output measures such as student test scores and examine the school context. Whether the indicators focus on inputs or outputs, care must be taken to protect the technical quality of the data, to create a level of analysis that isolates within-school differences, and to ensure that across-school comparisons are fair. A list of 29 references is appended. (FMW)

ED334339

Creating Accountability in Big City School Systems

Linda Darling-Hammond
Carol Ascher

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

NCREST
NATIONAL CENTER FOR RESTRUCTURING EDUCATION, SCHOOLS AND TEACHING

ERIC®
Clearinghouse on Urban Education

UD028189

BEST COPY AVAILABLE

CREATING ACCOUNTABILITY IN BIG CITY SCHOOLS

CREATING ACCOUNTABILITY IN BIG CITY SCHOOLS

Linda Darling-Hammond and Carol Ascher
Teachers College

Urban Diversity Series No. 102

ERIC Clearinghouse on Urban Education/
Institute on Urban and Minority Education

National Center for Restructuring
Education, Schools and Teaching

March 1991

ERIC Clearinghouse on Urban Education
Box 40, Teachers College
Columbia University
New York, New York 10027
212/678-3433

National Center for Restructuring
Education, Schools and Teaching
Box 110, Teachers College
Columbia University
New York, New York 10027
212/678-3432

Director: Erwin Flaxman
Managing Editor: Wendy Schwartz

Co-Directors: Linda Darling-Hammond
Ann Lieberman

This publication was produced by the ERIC Clearinghouse of Urban Education with funding from the Office of Educational Research and Improvement, United States Department of Education, under contract number R188062013, and from Teachers College, Columbia University; and by the National Center for Restructuring Education, Schools and Teaching with funding from the Dewitt Wallace-Reader's Digest Fund, Aaron Diamond Foundation, Leon Lowenstein Foundation and the New York Community Trust. The opinions expressed in this publication do not necessarily reflect the position or policies of the funders.

Copies are available from the ERIC Clearinghouse on Urban Education and the National Center on Restructuring Education, Schools and Teaching.

Copies are also available from the ERIC Document Reproduction Service (EDRS), 7420 Fullerton Road, Springfield, VA 22153, 1-800-443-ERIC, both on microfiche and paper. Contact the Clearinghouse or EDRS for full ordering information.

TABLE OF CONTENTS

DETERMINING WHO IS RESPONSIBLE FOR WHAT AND TO WHOM 1
- The Nature of Accountability 1
- Accountability as a System 2
- Types of Educational Accountability 3
- Bureaucratic Accountability 4
- Professional Accountability 5
- The Balance between Equity and Productivity 7
- Market Accountability 8
- School-Based Management 9
- Conclusion 11

USING STATISTICAL INDICATORS 13
- The Value of Indicators 13
- Criteria for Choosing Indicators 15
- Assessing School Context 16
- Criteria for Interpreting Indicators 17
- Indicators and Policy 22
- Who Should Choose the Indicators? 23
- Conclusion 24

CREATING ACCOUNTABILITY AT THE SCHOOL LEVEL 26
- The Role of Testing 26
- Creating an Accountability System 27
- Working with Different Levels of Accountability 28
- Creating Processes that Enhance School Accountability 29
- Creating Indices That Enhance School Accountability 32
- Going within and beyond the School as a Unit 36
- Indicators as Incentives 37

REFERENCES 38

DETERMINING WHO IS RESPONSIBLE FOR WHAT AND TO WHOM

The Nature of Accountability

Accountability has always been a basic concept in public education, as it is in all public affairs, although ideas about how to accomplish it have changed. In education, accountability implies not only that teachers, principals, and other school people should be held responsible to parents, other citizens, and their elected representatives, but that parents are responsible for sending their children to school.

In recent years, problems in urban schooling have given rise to the idea that carefully created systems of accountability might spur school improvement, and even school restructuring (McDonnell, 1989). Obviously, the "natural accountability" of face-to-face contacts in small communities is impossible in big city schools. Urban school districts are also often extremely large, and they serve many and diverse constituents and function with layers of bureaucracy. The goal is, therefore, to create planned structures for such schools and districts that can ensure accountability equal to or better than the "natural" kind.

As has quickly become clear, devising a system of genuine accountability is a complex task, involving a careful sorting of responsibilities and a thoughtful set of measures for assessing school effectiveness and student progress. Each attempt to make a big city school accountable must involve the questions: "Who is responsible?" "For what?" and "To whom?" (OERI State Accountability Study Group, 1989). The answers to these questions are not always easy to ascertain, and developing a system of accountability based on them is likely to be multifaceted and complex.

Even when the goal is to make teachers responsible for the quality of learning, certain methods of trying to achieve it constrain school practices in ways that actually make learning *less* likely for students. Given such complications, it is not surprising that policymakers have often reached for the apparently easiest strategy—monitoring students' test scores—and have sometimes even linked key decisions (promotions, graduation, teacher and school rewards or sanctions) to such measures, presumably as a means of achieving accountability.

However, performance indicators, such as test scores and dropout rates, are information for the accountability system; they are not the system itself. Accountability (i.e., responsible practice and responsiveness to clients) occurs only when a useful set of processes exists for interpreting and acting on the information. In fact, if school indicators are improperly designed or unintelligently used, they can actually undermine accountability.

Accountability as a System

An accountability system is a set of commitments, policies, and practices that are designed to:

1) heighten the probability that students will be exposed to good instructional practices in a supportive learning environment;

2) reduce the likelihood that harmful practices will be employed; and

3) provide internal self-correctives in the system to identify, diagnose, and changes courses of action that are harmful or ineffective.

Assessment data are helpful to the extent that they provide relevant, valid, timely, and useful information about how much individual students are learning and how well schools are serving them. But these kinds of data are only a tiny part of the total accountability process. Accountability also encompasses how a school or school system hires, evaluates, and supports its staff; how it relates to students and parents; how it manages its daily affairs; how it makes decisions; how it ensures that the best available knowledge will be acquired and used; how it evaluates its own functioning, as well as students' progress; how it tackles problems; and how it provides incentives for continual improvement.

Types of Educational Accountability

In education, as in other enterprises in our society, at least five types of accountability mechanisms exist along side each other (Darling-Hammond, 1989):

- *Political Accountability*. Legislators and school board members, for example, must regularly stand for election.
- *Legal Accountability*. Citizens can ask the courts to hear complaints about the public schools' violation of laws regarding, for example, desegregation or equal educational opportunity.
- *Bureaucratic Accountability*. District and state education offices promulgate rules and regulations intended to ensure that schooling takes place according to set standards.
- *Professional Accountability*. Teachers and other school staff must acquire specialized knowledge, pass certification exams, and uphold professional standards of practice.
- *Market Accountability*. Parents and students may choose the courses or schools they believe are most appropriate. They may also be involved in other, more direct means of participating in school decision-making.

All of these accountability mechanisms have their strengths and weaknesses, and each is more or less appropriate for certain goals. Political mechanisms can help establish general policy directions, but they do not allow citizens to judge each decision by elected officials and they do not necessarily secure the rights of minorities. Legal recourse is useful in establishing and defending individual and group rights, but not every education decision is subject to court action and not all citizens have access to the courts.

Bureaucratic mechanisms are appropriate when standard procedures will produce desired outcomes, but they can be counterproductive when clients have unique and different needs. Professional mechanisms are important when services require complex knowledge and decisionmaking to meet clients' individual needs, but they do not always take competing public goals, such as better services versus cost containment, into account. Finally, market mechanisms are helpful when consumer preferences vary widely and the state has no direct interest in controlling choice, but they do not ensure that all citizens will have access to services of a given quality.

Because of these limits, no single form of accountability operates alone in any major areas of public life. The choice of accountability tools, and the balance among different forms of accountability, is constantly shifting as problems emerge, as social goals change, and as new circumstances arise.

In education, the power of electoral accountability, exemplified in the authority of school boards, has waxed and waned over the past 20 years. In the same period, legal forms of accountability have grown, as court cases have been used to settle educational policy, and bureaucratic forms of accountability have expanded through increased policymaking at the district and state levels. Recently, market accountability, the least used form, has been expanded somewhat through magnet schools and school choice. Finally, based on an expanding knowledge base and efforts to create more meaningful standards of practice, professional accountability is gaining currency as a way to improve teaching. Nevertheless, of all these forms, bureaucratic and legal accountability now outweigh all others—and some experts suggest that these two forms may have overextended their reach.

Because bureaucratic, professional, and market accountability tools are all currently being proposed as strategies for school improvement, it is important to understand what each offers and where they may conflict.

Bureaucratic Accountability

The strength of bureaucratic accountability is its goal of ensuring equal and standardized education. In a bureaucratic approach to school management, schools are seen as agents of government to be administered by hierarchical decisionmaking. Policies are made at the top of the system and handed down to administrators who translate them into rules and procedures. Teachers follow these rules and procedures (class schedules, curriculum guides, textbooks, rules for promotion and assignment of students, etc.), and students are processed according to them.

Bureaucratic accountability offers the hope of finding the "one best system," codified by law and specified by regulations, under which all students may be educated. Bureaucratic forms of accountability assume the following:

1) that students are sufficiently standardized so they will respond in identical and predictable ways to the "treatments" devised by policymakers and their principle agents;

2) that sufficient knowledge of which treatments to prescribe is both available and generalizable to all educational circumstances;

3) that this knowledge can be translated into standardized rules for practice, which in turn can be maintained through regulatory and inspection systems; and

4) that administrators and teachers can and will faithfully implement the prescriptions for practice.

In the bureaucratic model, teachers are held accountable for implementing curricular and testing policies, most often prescribed at the district and state levels, whether or not the prescriptions are appropriate in particular instances for particular students. Their own knowledge about learning theory and pedagogy may actually be a liability if it conflicts with these policies. (In fact, current knowledge about learning suggests that such conflicts will frequently occur, since students learn in different ways and at different rates under different circumstances, and effective teaching strategies must be flexible, adaptive, and non-uniform to meet these needs.) However, the bureaucratic model does not seek to ensure that teachers are highly knowledgeable, since it presumes that many important decisions will be made by others in the hierarchy and handed down to teachers in packaged form. Under this system, teachers cannot be held accountable for meeting the individual needs of their students; they can only be held accountable for following standard operating procedures.

Insofar as bureaucratic accountability mandates standardized curricular and instructional practices, the options for school-based management (which gives meaning to magnets and other schools of choice) are also limited. Thus, both professional accountability and market accountability may conflict with certain forms of bureaucratic accountability.

Professional Accountability

Professional accountability, whether for doctors, lawyers, or teachers, starts from very different assumptions than does bureaucratic accountability. Unlike bureaucratic accountability, whose goal is uniformity and standardization, professional accountability seeks to create practices that are client-oriented and knowledge-based.

It assumes that, since decisions about different clients' needs are too complex and individualistic to be prescribed from afar, the system must be structured to ensure that practitioners will be able to make responsible decisions. It aims, therefore, to ensure professional competence through rigorous preparation, certification, selection and evaluation, and continuous peer review. It requires that practitioners make decisions on the basis of the best available professional knowledge, and that their first commitment be to the welfare of the client.

Rather than aiming to regulate practices, this model seeks to ensure that the persons charged with students' care will be highly knowledgeable, competent, and committed, as parents, and students themselves, have a right to expect. The quality of staff hired and retained is a key component of professional accountability. Consequently, a professional accountability system must pay particular attention to personnel policies governing the preparation, hiring, and evaluation of teachers and other staff; the support given to their ongoing learning; and the assessment vehicles that exist for evaluating classroom and school practices, as well as student progress.

For educators to make the intellectual and ethical decisions inherent in professional accountability, they must have the knowledge and skills to support responsible decisionmaking. They must also work under fewer rules about what is taught, when, and how, so that decisions can be made in response to students' needs rather than in response to regulations.

Because professional responsibility is explicitly directed toward the diverse needs of students, it often contravenes standardized practice, and thus some forms of bureaucratic accountability. A professionally responsible teacher may decide that some or all of his or her students would be better taught using books other than a mandated text that is poorly constructed, at an inappropriate reading level, or biased in its depiction of certain racial, ethnic, or cultural groups. The teacher may be aware that the learning styles of some students would be better addressed by one set of teaching methods or materials than by another prescribed for general use, and would insist, drawing on professional knowledge about assessment, that no decision about any student be made solely on the basis of a single test score or other standardized measure.

The potential benefit of professional accountability is a focus on the appropriateness of decisions for students' needs. The potential pitfall is that, if professional standards are not rigorously enforced for preparation and for ongoing practice, educators' individually-oriented decisions may become idiosyncratic rather than well-grounded in professional knowledge and ethics.

The Balance between Equity and Productivity

In deciding which aspects of education should be relegated to bureaucratic accountability and which should be left to professional accountability, it is useful to distinguish between equity and productivity concerns (Wise & Gendler, 1989). Equity issues generally must be resolved by higher units of governance, because they "arise out of the conflicting interests of majorities and minorities and of the powerful and powerless," and because local institutions are often captive to majoritarian politics and "intentionally and unintentionally discriminate" (p. 206). Matters such as the allocation of resources and guarantees of equal access can and should be regulated by a higher level of government.

On the other hand, productivity questions cannot be solved most effectively by bureaucratic regulation, since, at its best, teaching knowledge is used in a highly individualized manner. Thus, state- and district-level policy decisions about teaching methods and school processes cannot meet the needs of varying school and student circumstances. Improving student and school achievement demands discretionary decisions safeguarded by professional accountability.

The growth of bureaucratic forms of accountability in recent decades has made it clear that top-down decisionmaking cannot solve all problems, and that overregulation may at times undermine accountability. In addition to prescribing practices that are likely to be inappropriate for some students, hierarchical decisionmaking often leaves no one accountable for results. When school level staff are not responsible for making decisions, they also do not become responsible for finding solutions to school problems.

The fact that different student needs and school circumstances require different strategies has brought renewed attention to methods for lodging greater authority and responsibility at the school and classroom level. Two concepts currently receiving a

great deal of attention are school choice, which aims to create more market accountability, and school-based management, which relies on greater professional and political accountability.

Market Accountability

Magnet schools and other choice plans are based on quasi-market mechanisms. These market mechanisms are supposed to make schools more accountable in at least two ways: (1) by letting "customers" choose, schools are expected to work harder to provide services that parents or students want; and (2) by allowing choice, the market is expected to reveal that undersubscribed schools have problems that policymakers should address.

Several questions are raised by market accountability as it now exists. Most immediately are these: How do students and their parents currently make school choices? What indicators should schools provide to the public for students and their parents to be informed consumers of public education? What positive measures, other than changing schools, can parents and students take when schools do not meet their needs? And, what vehicles might increase feedback about parents' and students' opinions of schooling? The possibility of greater feedback to schools and policymakers, and greater options for informed consumers, are among the major potential benefits of choice plans.

Even under the best conditions, however, choice alone is not enough to ensure full client accountability. For one thing, when the most desirable schools have already been filled, there are still a great many students left to be served and a number of schools, desirable or not, which must serve them. If no other public policy mechanisms are in place to support improvements in all schools, choice will only make marginal adjustments in the mix of students in some schools; it will not produce a general increase in the quality of education across the board.

Not surprisingly, market accountability can create tensions with other forms of accountability. Insofar as bureaucratic accountability standardizes what is offered in different schools, it minimizes meaningful choice. Even though professional accountability implies a client orientation, parent participation may be problematic for

teachers, since decisions based on professional knowledge may not always coincide with parents' wishes.

Finally, accountability in public education has traditionally assumed that the state, district, and school should all be held accountable at different levels for public education: the state for equitable financing, the district for hiring practices and other resource allocations, and the schools for curriculum and teaching decisions. But in cases of interdistrict choice, where students can attend any school in a state, the district loses its clear-cut public. As voters, parents hold the state accountable; as consumers choosing a particular school, they can hold the school accountable. But once their children leave the district in which they reside, they sever electoral accountability at the district level.

Other proposals, such as those for school restructuring and new forms of school management, suggest that client accountability must also be promoted through school structures for shared governance, accessible review and appeals processes, and parent involvement in decisions about their own children.

School-Based Management

In the last several years, as emerging research has suggested that bureaucratic accountability, with its emphasis on regulatory standardization, is counterproductive for creating the best education for diverse bodies of students, school restructuring proposals have called for local ownership of education, or school-based management. These proposals assume that the greatest accountability is owed to students and their parents, and that teachers and other professionals at the school site should be predominantly accountable.

Most school-based management proposals call for shared decisionmaking among faculty, staff, parents, and students. They assume that better decisions will be made when those who are closest to the situation, and who must live with the decisions, are involved. Where parent and student participation is called for, these initiatives introduce a form of electoral accountability that is otherwise very weak in large, impersonal school systems. As greater authority is vested in school faculties, professional accountability mechanisms, which aim to ensure the competence and

commitment of staff, should also be strengthened so that this authority will be well-used.

Several questions arise: How much latitude should be left to school level communities to define and monitor their educational processes and outcomes? Conversely, in a system where significant authority devolves to the school level, how much responsibility should state governments and local districts still bear for students' treatment and for their achievement? A key question for school-based management proposals is whether states and districts should be accountable for regulating inputs (resources) or outcomes (measures of achievement).

Wise and Gendler (1989) have argued persuasively that, when a state guarantees equality of financial resources, it encourages local initiative, equalizes the capacity of poor districts to secure a sufficient and highly qualified teaching force, and permits schools from poor districts to choose among curriculum and equipment options, just as wealthy districts do. When a state regulates outputs, however, by mandating achievement outcomes defined by standardized tests, its "effort to produce equal education ends up degrading learning for all. Individuality, creativity, and depth are lost; all that is retained is uniformity, conventionality, and trivial skills" (p. 36).

Yet, states must have some way of evaluating how well schools are discharging their responsibilities to students. In some cases, this evaluation function may be best delegated to local school districts, accreditation agencies, or other organizations with special expertise. Regardless of who evaluates, however, a major problem is that the measures used to evaluate student and school progress—and the ways in which these measures are used in decisionmaking—have major effects on the quality of education itself. Thus, it is important to decide who chooses the measures for assessing students and schools, and whether those measures are valid and useful for the different purposes they may be asked to serve.

A related concern is the balancing of different accountability mechanisms as changes are sought. If schools are to rely less on hierarchical regulation to define their processes, then other forms of accountability must be strengthened to protect students' welfare. Greater guarantees of staff competence and commitment would accompany more stringent professional accountability mechanisms. Greater voice for parents and

students would accompany more powerful political accountability mechanisms involving participatory decisionmaking at the school site. Greater incentives to attend to consumers' wishes might result from enhanced market accountability mechanisms involving school choice. The key is to find the right mix of tools to provide support for school improvements that will encourage responsible and responsive education.

Ideally, each level of the system would assume its appropriate share of responsibility:

- *States* would be responsible for providing equal and adequate resources to schools and for ensuring the enforcement of equity standards and standards of professional certification.
- *School districts* would be accountable for the policies they adopt (including everything from staff hiring standards to paperwork requirements), for equity in the distribution of school resources, and for creating processes that make them responsive to the needs and concerns of parents, students, and school-level staff.
- *Schools* would be accountable for equity in the internal distribution of resources, for adopting policies that reflect professional knowledge, for establishing means for continual staff learning, for creating problem-identification and problem-solving processes that drive continual improvements, and for responding to parent, student, and staff ideas.
- *Teachers* would be accountable for identifying and meeting the needs of individual students based on professional knowledge and standards of practice, for continually evaluating their own and their colleagues' practices, for seeking new knowledge, and continually revising their strategies to better meet the needs of students.

Conclusion

Massive testing or any other data collection effort does not create an accountability system, nor does it guarantee improvement in urban or non-urban schools. A school or school district creates various policies and practices that make it more accountable by using many different tools, including methods for teacher and parent participation in decisionmaking, bureaucratic regulations, legal recourse, safeguards and support for the competence of staff, and options for choice. Data about student and school progress should inform the system so that responsible decisions are made and problems are corrected when they arise.

Since each accountability tool has different strengths and weaknesses and provides different incentives, a careful blend of methods is needed to improve schools for all students. Although simple answers are appealing, it is only by struggling with tough questions of who is responsible, for what, and to whom that education, particularly in large urban schools, becomes truly responsive to students, parents, educators, politicians, and the general public.

USING STATISTICAL INDICATORS

The Value of Indicators

Recently, in response to pressure to improve American education, particularly in urban areas, states, districts, and local schools have moved rapidly to put accountability systems into place. Although there is widespread pressure on schools to find better ways to attend to their "bottom line"—student learning—ideas about how to both stimulate and measure school improvement are still in their infancy. One aspect of the accountability quest is the search for types of information about school and student performance that can be used on a regular basis to inform policymakers, the public, and educators about educational trends and needs. While most educators and researchers agree that accountability systems need multiple indicators, there is less agreement about what these indicators should be, or who should be responsible for determining them.

There is also disagreement about how indicators should be used in an accountability system. Some proposals suggest that measures of educational outcomes, such as standardized test scores or school graduation rates, can be used to "hold schools accountable" by triggering rewards, sanctions, or remedial actions. The implication is that accountability is achieved by collecting a limited number of performance statistics and using them to "motivate" schools.

Other views suggest that a wider range of indicators is probably necessary for understanding how schools are performing and whether students are benefiting from their studies. This is because it is difficult to interpret outcome measures without also examining information about student characteristics, along with school processes,

activities, and resources. Increases or declines in test scores, for example, may be due to school variables—changes in school practices, teacher quality, or teaching methods—but they may also be the result of many other factors, including changes in the population of students taking the test.

Furthermore, there is the question of whether and how the availability of statistical data makes schools more responsible for using good practices and more responsive to the needs of their clients, that is, more accountable. As noted in the previous section, accountable schools establish policies and practices in all areas likely to produce responsible and responsive education for their students. Their accountability systems enable them to set goals or standards and to diagnose how well they are meeting them; the schools establish processes by which these standards are likely to be achieved or maintained; and they include ways of correcting problems when they arise. While good indicators can help inform and improve school operations, the indicators themselves do not create accountability. In fact, indicators that are poor measures, or are badly used, can actually undermine accountability by creating incorrect assumptions about what schools are doing and what students are learning.

Whatever the chosen indicators, each is generally used to support broad inferences—for instance, whether there has been a decline in educational achievement, or how U.S. students' mathematics achievement fares in comparison to those in other countries. At the same time, because of cost, a small number of indicators tend to be chosen because they are simple to gather. Concerns about cost and convenience can easily result in failure to assess higher order thinking skills well, since such assessment requires tasks that take students longer to perform and costs more to evaluate. Similar concerns for cost and efficiency may also prevent collection of enough data to test alternate hypotheses for the trends observed. For example, are average test score increases due to more advanced course taking by some students, or due to increased student dropout rates by low-scoring students? In the latter circumstance, an increase in average test scores, caused by exclusion of the lower scorers, might occur without any student actually increasing his or her achievement.

Clearly, it is important for those developing indicator systems to consider carefully *what* they want to measure, *how* they can best measure it, and *what other information* they need to interpret trends intelligently. As individual schools, school

districts, and State Education Departments begin to develop more complex accountability systems, several issues concerning indicators should be considered.

Criteria for Choosing Indicators

Indicators should offer at least one of the following types of information (Oakes, 1986; Koretz, 1989):

1. *Problem-oriented information.* Indicators should be able to detect current problems or alert school systems to potential difficulties. Tracking teacher turnover or student mobility are examples.

2. *Policy relevant information.* Indicators should not only describe conditions of interest to policymakers, but the conditions should be described in a way that is amenable to change by policy decisions. For example, if state policymakers want more students to take academic courses in high school, indicators of course taking should show which types of students in which types of schools or districts are not taking these courses, so that policymakers will know where to target their efforts.

3. *Information on educational outcomes.* Data might include graduation or dropout rates; college attendance rates; measures of responsible citizenry, such as voting rates or participation in civic service activities; or measures of learning (e.g., achievement test scores; assessment of student work products, such as writing samples; evaluation of participation in exhibitions such as science fairs, student publications or newspapers, debates, dramatic productions).

4. *Information on students' backgrounds and placements.* Indicators cannot be interpreted meaningfully without knowing which students are involved. For example, if a school's student population is highly mobile, test scores intended to measure student growth at different points in time may not even be measuring the same students. Similarly, monitoring equality of access to school opportunities requires knowing which school outcomes are associated with different groups of students and what types of schooling they are experiencing.

5. *Information about school context factors.* Some school context data, such as the qualification of teachers, or the use of teaching methods known to encourage the acquisition of higher order skills, have already been shown to be linked to outcomes. However, other data might provide information on central inputs and processes of the system—for example, financial resources, teachers' work load, schools' general curriculum offerings. These are needed in order to understand how the system works, even though research may not as yet have linked all of them directly to performance.

Assessing School Context

In order to understand what is happening in schools, it is important to have indicators to characterize the resources, people, and activities that shape students' school experiences. Oakes (1989) notes that school context indicators can offer clues about why schools achieve the outcomes they do, and they can prevent schools from de-emphasizing important but unmeasured goals in order to "look good" on the limited outcome measures that are now available. School context measures enable those using an accountability system to disaggregate data by gender, race, and class, and to examine the availability of equal educational opportunities. They also allow disaggregation by important subgroups, such as students in academic or vocational tracks.

Although fiscal and other school resources (relatively easy to tabulate) are a necessary prerequisite for good schooling, they don't directly correlate with student achievement; this is because such organizational practices as student placement policies, curriculum decisions, and the time devoted to learning, all mediate the use of resources (Madaus, Kellaghan, Rakow, & King, 1979). Nonetheless, resources do influence students' opportunities to learn and thus ultimately affect their achievement. Therefore, it is important is to understand where various resources fit in the larger scheme of schooling so that useful measures can be constructed.

Drawing on a large base of educational research, Oakes (1989) suggests that three interrelated variables are important to school achievement, and that each of these can be tapped with a set of indicators. The first, *access to knowledge,* is the extent to which schools provide students with opportunities to learn various domains of knowledge and skills. The second, *press for achievement*, is the mix of opportunities and incentives schools provide students to work hard and achieve. The third, *professional teaching conditions*, are the environmental factors that can empower or constrain teachers and administrators as they create and implement instructional programs. As shown on the chart below, a set of indicators can be created for each.

INDICATORS OF SCHOOL CONTEXT

Access to Knowledge	Press for Achievement	Teaching Conditions
Teacher Qualifications	Administrative Involvement in Academics	Teachers' Salaries
Instructional Time	Graduation Requirements	Pupil Load/Class Size
Course Offerings	Enrollment in Rigorous Programs	Teacher Time for Planning
Grouping Practices	Recognition of Academic Accomplishments	Collegial Work
Materials, Equipment	Academic Expectations	Involvement in Decisions
Teaching Methods	Quality/Type of Homework	Evaluations Emphasizing Learning
Academic Supports	Uninterrupted Class Instruction	Teacher Flexibility
Enrichment Activities		Administrative Support for Innovation
Parent Involvement		Clerical Support
Faculty Beliefs		

These are just samples of the kinds of school context factors that influence school performance and student learning. Clearly, it takes some ingenuity to find appropriate indicators for such school features as "teaching conditions" and "press for achievement." Moreover, no indicator will be useful alone; a comprehensive group must be drawn from within each of several categories if the kind and quality of education provided by a school or school system is to be seriously judged.

Understanding the potential benefits of a more comprehensive indicators system, some states and districts have begun to collect such data as the level and use of fiscal resources, qualifications of the teaching staff, courses offered, and students' course taking patterns, as well as promotion and dropout rates of students and information about students' backgrounds (McDonnell & Oakes, 1988).

Criteria for Interpreting Indicators

One of the greatest dangers of indicators is the ease with which they can be misunderstood—most often because they are used or interpreted in invalid ways, thus giving false impressions. Some of the most common misinterpretations are caused by using average test scores to indicate the quality of schools or school districts.

Most people understand that, since schools serve students with very different educational needs and starting points, student average test scores alone are not a valid measure of the quality of instruction offered. However, many do not understand that even changes in a school's scores from one year to the next cannot capture the effects

of school quality. The following mini-sample of two years of test scores from Lincoln Junior High School suggests how changes in average scores can misrepresent true student achievement.

Lincoln J.H.S. Test Scores

1987	1988
10	9
9	8
8	7
7	6
2	
36	30
Mean = 7.2	7.5

The numbers represent individual student scores on a criterion-referenced test given in two successive years. Although the increase in mean test scores (from 7.2 to 7.5) would suggest that the school is improving, in fact, the scores of all of its students declined. The illusory improvement is due solely to the fact that the lowest-scoring student is no longer listed, most likely because he transferred, dropped out, or was placed in special education so his test scores wouldn't be counted. In fact, where this type of measure is used to evaluate schools, especially when "high stakes," such as rewards or sanctions, are attached to test scores, incentives are created for pushing low scoring students out of the school, or at least out of the test score count (Darling-Hammond, 1990a; Haney & Madaus, 1986).

This example illustrates one problem of validity. (Validity and reliability are two of the technical features that are relevant in choosing and using an indicator.)

Validity does not inhere in the measure itself, but refers to the relationship between the measure and the inferences drawn from it. If averages hide information about individual students, then interpretations about school or student progress drawn from indicators based on trends in average test scores will be invalid. Similarly, teachers' "teaching to a test" can invalidate the inferences made about test score meanings. This is because test items are only meant to be samples of domains of

learning, and the assumption that student performance on those items fairly represents the broader domain of curriculum goals is violated when the focus of teaching is on only those items to appear on the test (Koretz, 1988).

Reliability refers to the degree to which an indicator is free of random errors of measurement. That is, whether a student's score on a test is likely to change significantly from one administration to another; or whether statistics on a school's curricular offerings are likely to vary depending on who is collecting the data.

The problems of validity and reliability raise a number of related concerns which threaten the accurate interpretation of indicators.

Corruptibility of Indicators. Unfortunately, as argued above, it is possible to change a measurement without causing a comparable change in the construct it is supposed to measure. For example, the emphasis on basic skills test scores in reading and mathematics has prompted teaching to the test, which can mean that the scores no longer necessarily indicate students' general achievement, because students spend less time on other subjects such as science, social studies, or the arts (Darling-Hammond & Wise, 1985). It can also mean that the scores are no longer comprehensive assessments of students' ability even in the tested subjects. As classwork becomes oriented narrowly toward the test, aspects of the subjects that are not tested, such as analysis, complex problem-solving, and written or oral expression, are left untaught.

Not surprisingly, since about 1970, while U.S. students' basic skills test scores have been increasing, scores on assessments of higher order thinking skills have been steadily declining in virtually all subject areas. Officials of the National Assessment of Educational Progress, the National Research Council, and the National Councils of Teachers of English and Mathematics, among others, have all attributed this decline to schools' emphasis on tests of basic skills.

As the National Assessment of Educational Progress (NAEP) found: "Only 5 to 10 percent of students can move beyond initial readings of a test; most seem genuinely puzzled at requests to explain or defend their points of view." The NAEP assessors reported that current methods of teaching and testing reading require short responses and lower level cognitive thinking, resulting in "an emphasis on shallow and superficial

opinions at the expense of reasoned and disciplined thought..., [thus] it is not surprising that students fail to develop more comprehensive thinking and analytic skills" (NAEP, 1981).

Test scores are not the only indicators that can be corrupted. For example, indicators of student course taking may prove to mean something other than what policymakers intended if the courses are relabeled or modified to give them the appearance of meeting requirements. When pressure is great to meet new standards, schools with poor resources or shortages of qualified teachers may implement new mathematics or science requirements quite differently from wealthy schools with high achieving students. In one case, three years of science courses may mean three years of warmed-over but reclassified general science lectures; in the other, it may mean a full and rigorous sequence of biology, chemistry, and physics classes replete with laboratory experiences. When an indicator has different meanings in different settings, it suffers from problems of both reliability and validity, and cannot be interpreted as intended.

Clearly, for an indicator system to work, everyone participating in the system—even those who are only supplying the data—must perceive it as serving their needs and interests. To the extent that an indicator system either poses a threat, for example, to student promotion, school funding, or staff job security, or offers the possibility of a reward, the information fed into it will likely be corrupted. To the extent that the indicators in a high stakes situation are only partial measures of the real goals, so too are the unmeasured goals of the system likely to be sacrificed (Haney & Madaus, 1986).

Scope of the Indicator System. Every indicator system necessarily has its focus or foci. However, indicators should also be broad enough to catch unintended side effects. Otherwise, the ways in which policies affect the educational system may go unnoticed, and incorrect inferences may be drawn about what is actually occurring (Koretz, 1989).

For instance, if a state or district does not have enough foreign language teachers, a policy intended to increase enrollments in introductory Spanish can decrease the number of advanced Spanish language courses offered as teachers are deflected

from one placement to another. Similarly, because there is limited room in the curriculum unless time for schooling is increased, when only some aspects are regulated, more emphasis on one subject may mean less on another. Increasing science requirements may result in students' taking fewer history courses. Raising graduation requirements may stimulate more course taking for some students but increase the chance of dropping out for others.

If an indicator system is comprehensive, it will include measures of the many aspects of the educational system: school resources and offerings, teacher availability and qualifications, student participation and performance. This allows unintended side effects to be monitored and appropriate changes to be made in order to pursue the genuine goals of the policies.

Making Fair Comparisons. Accountability systems often involve comparing schools and school districts. Moreover, rewards and punishments are increasingly associated with doing well or poorly on these comparisons. Thus, it is important that comparisons are set up fairly. This means that school outcomes cannot be compared without also comparing populations (and population shifts over time), resources, and even educational goals.

One way to create fair comparisons is to develop longitudinal measures of student progress—criterion-referenced measures of how much each pupil has learned over a period of time—as a basis for aggregating the gain scores for students by schools. This approach, used on a limited basis in some New York City schools, more accurately examines student progress than does the standard practice of comparing the scores of schools with very different student populations, or even the practice of looking at changes in average test scores (say, for two different groups of third grade students at two different points in time), particularly in schools with high levels of student mobility.

Another strategy is to compare schools with "like" schools. California, for example, provides each school with a detailed report comparing it to schools with similar student bodies on a wide range of indicators. These indicators include students' participation rates in academic courses, attendance and dropout rates, social attitudes

and perceptions regarding the school; and mobility data and test score information disaggregated by income group (Archbald & Newmann, 1988).

The California approach, although not perfect, is a marked improvement over a statistical strategy that has sometimes been used to "adjust" schools' test scores, based on student characteristics. Unfortunately, these technical "adjustment" formulas reify the effects of discriminatory schooling practices by obscuring differences in resources in low-income and minority schools that should be highlighted if comparisons in the service of genuine accountability are to be made.

An example of how comparative indicators may be used to obscure or illuminate factors influencing achievement can be seen in the case of mathematics learning. Data from many sources have long shown differential achievement between non-Asian minority students and white students in mathematics. Yet, comparative indicators have only recently demonstrated that mathematics course taking—which is strongly related to achievement test scores—varies substantially by racial/ethnic group. In fact, for students with similar course taking records, achievement test score differences by race or ethnicity virtually disappear (Oakes, 1989).

The inequalities in course taking are in large part due to the fact that schools serving minority, low-income, and inner-city students have far fewer qualified teachers and offer many fewer advanced courses than more advantaged schools do (Oakes, Ormseth, Bell, & Camp, 1990; Darling-Hammond, 1990b). Comparisons of test scores that ignore these factors hold little promise of directing policymakers' attention to rectifying the real sources of the problem.

Indicators and Policy

Although indicators don't provide policy text, they do provide policy fodder, and they are powerful political tools. On one hand, indicators can generate public opinion in favor of policies that may be educationally beneficial—such as recruiting a greater supply of well-qualified mathematics teachers in order to support a richer mathematics curriculum for all students. On the other, they may indirectly persuade

the public to endorse policies that are educationally harmful—such as using test scores as the basis for student promotion.[1]

More directly, indicators are increasingly being used to evaluate teachers and judge schools, and in general as an "anxiety factor" to generate school improvement. Some of the dysfunctional consequences of misusing specific measures as tools for prodding school change were noted above. Where rewards and sanctions are automatically triggered by increases or decreases on limited measures, the perverse side-effects are still more dangerous, since students who perform poorly may be pushed out and talented staff may reject assignments at schools serving children with more challenging educational needs.

Under these conditions, indicators can be seen as undermining, rather than enhancing, genuine accountability, since they are used to shirk responsibility for careful analysis and complex decisionmaking in favor of simplistic, though damaging, "cures." If policymakers and educators are to be truly accountable for serving students in responsible and responsive ways, they must use information about educational conditions along with knowledge of sound educational practices to evaluate where changes are warranted and to adopt strategies that will support student success.

Ultimately, indicators should be seen as one set of inputs into a process of "reflective policymaking" and as grist for a related process of reflective school decisionmaking. They should help identify areas where further examination is called for and provide clues to promising lines of effort.

Who Should Choose the Indicators?

Any education accountability system contains not only indicators, but mechanisms which make certain individuals (teachers, principals, etc.) responsible to others. Until now, the tendency has been to build accountability systems top-down. Most recent indicator systems have been largely defined by states, and they have helped

---

[1] While a recent poll found that three-fourths of the public favored testing as a prerequisite for promotion and graduation (U.S. Department of Education, 1987), research demonstrates that students who are retained in grade ultimately learn less and are more likely to drop out than students of equal achievement levels who were promoted (Holmes & Matthews, 1984; Carnegie Council on Adolescent Development, 1989).

to consolidate decisionmaking authority at the state level. Often, school districts have had to shift curriculum priorities in order to succeed according to the states' indicator systems. The result has sometimes been a de-emphasis on types of learning that had been valued in their local communities. Thus, one of the strengths, as well as dangers, of an indicator system is that, at the same time as it may make education more accountable to those who choose the indicators, it can also decrease the responsiveness of the system to others.

In order to be responsive to local parents and to the individual needs of students, local schools often need to collect information that is quite different from that required by the state. They must collect information that is useful for deciding what is working or failing in a particular school, academic track, or classroom. More important, since research on school improvement demonstrates the importance of the school site as the locus for change (Oakes, 1989), locally developed indicators may prove to be more effective educationally (see section below).

Because there has been so much accountability activity at the state and national levels, it might be easy to forget that both the state and federal governments are only acting on behalf of parents and other citizens. Thus, regardless of where they are instituted, indicators should provide real and understandable information about the quality of education being offered. And, because parents may respond to indicator information by enrolling or taking their child out of a school, it is important that this information be as rich as possible—and certainly not misrepresent—what is going on in a particular school or school district.

Conclusion

Policymakers are just beginning to go beyond such rudimentary output indicators as students' test scores to understand school resources and activities. This wider notion of indicators is essential for any inferences about the meaning of educational outputs.

The development of school context indicators is particularly important, and identifying the appropriate formats for eliciting needed information takes some effort. An important point to keep in mind is that school context can only be elicited by

multiple indicators that simultaneously cover such arenas as access to knowledge, press for achievement, and professional teaching conditions.

Whether the indicators focus on inputs or outputs, care must be taken to protect the technical quality of the data, to create a level of analysis that isolates important within-school differences in learning opportunities, and to ensure that across-school comparisons are fair. It is also important to exercise continual vigilance about the corruptibility of indicators, so that they not be used in a short-sighted manner that obscures the realities of what schools are offering, and to whom.

While indicators aim to monitor the health or status of a school system in order to inform policy, their policy implications can easily be overblown. Indicators cannot provide the specifically designed and detailed information necessary to answer fundamental research questions or evaluate specific programs and policies, nor can they define what is good education or desirable educational legislation.

Finally, indicators are not a substitute for either educational ideas or decisions about which policies should be implemented. The latter will always be influenced by values as well as knowledge. As Oakes has written, "Indicators can not remove fundamental decisions about schooling from the political process—from the reflection and debate among policymakers and the public ultimately responsible for its healthy functioning" (1986, p. 23).

CREATING ACCOUNTABILITY AT THE SCHOOL LEVEL

The Role of Testing

In the quest for accountability there has been a tendency to look to testing as a major mode of measurement and a tool for reform. Unfortunately, testing does not ensure that schools will teach students well or responsibly, or that students will learn what their parents and the public want them to know. Two recent major studies of American education have called attention to the degradation of learning that has accompanied the growth of standardized testing. Ernest Boyer's study (1983) of American high schools found an overabundance of teaching consisting of the transmittal of "fragments of information, unexamined and unanalyzed." Boyer notes:

> The pressure is on to teach the skills that can be counted and reported. As one teacher said, "We are so hung up on reporting measured gains to the community on nationally normed tests that we ignore teaching those areas where it can't be done."

Similarly, John Goodlad (1984) found in his massive study of more than 1000 American classrooms that for the most part, students listen, read short sections in textbooks, respond briefly to questions, and take short-answer or multiple-choice quizzes. They rarely plan or initiate, create their own products, read or write anything substantial, or engage in analytic discussions. And there are few incentives for their teachers to pursue these approaches. As Goodlad comments:

> Teachers are sensitive to the pressures that state and district testing programs place on them. They get the message. The other messages—that there are goals beyond those that the tests measure, that pursuing such goals calls for alternative teaching strategies, that the fundamentals of the curriculum transcend grade-level

> requirements—are faint to begin with, and they are drowned out by the more immediate and stronger message...

These studies point out how important it is for schools to choose their "accountability tools" carefully. If performance measures are actually to support meaningful accountability, they must assess and encourage valuable kinds of learning for students. Furthermore, for such tools to be useful at the school level, they must illuminate how individual students are progressing, and they must be accompanied by methods for figuring out how changes in practice might be made to improve student learning.

Creating an Accountability System

As noted in the first section, an accountability system is a set of commitments, policies, and practices that are designed to create and nurture good practices and continual self-evaluation.

Thus, the ways in which schools hire, support, and evaluate staff are part of a genuine accountability system. So are the ways in which they structure teaching and learning; make decisions about curriculum; develop policies about student grouping, promotion, and discipline; create communication mechanisms between and among teachers, students, and parents to solve problems that arise; and develop means for evaluating how individual children and the school as a whole are performing.

Schools vary in the responsibility of their educational practice, as well as in their responsiveness to the needs and views of students, parents, and staff. For example, a school that exercises great care in hiring and supporting competent and committed staff is a more responsible school than one that either hires teachers without adequate consideration of their competence or places teachers in conditions that undermine their ability to teach competently. A school that provides many occasions for faculty to meet with each other and with parents to evaluate and discuss their teaching and students' progress is a more responsive school than one that provides no opportunities for staff and parents to engage in problem-solving to promote improved teaching strategies. A school that ensures that students receive individual attention to their needs (and actively seeks out knowledge to meet those needs) is a more responsible school than one that treats students en masse, allowing many to fail.

The point is that accountability is achieved only if a school's policies and practices work both to provide good education and to correct problems as they occur. Both *performance indicators* and ongoing *diagnostic processes* are needed to evaluate whether these conditions are being met. These processes include occasions to assess how the school is functioning and how well individual student needs are being met. They also include methods for changing school practices—even restructuring certain aspects of schooling—if they are not working well on behalf of students.

Working with Different Levels of Accountability

All schools are different. No top-down management structure can fully serve the unique needs of teachers and students in a particular setting (Lieberman & Miller, 1990). While centrally-collected and analyzed information can sometimes be useful to a school in identifying issues for further consideration, the school's improvement strategies must be based on local goals, issues, problems, and change processes, as well as on more fine-grained, immediately relevant information. To recognize the unique qualities and to analyze the special challenges of a school, each school must design its own accountability structures.

Unfortunately, top-down accountability systems that consume a good deal of a school's time and resources can sometimes discourage individual schools from working to achieve their own accountability systems. Nonetheless, if schools do not work themselves to create more responsible and responsive practices, others outside the school are certain to fill the void with even more intrusive approaches that are less likely to demonstrate sensitivity to the school context or a full appreciation of the school's strengths and human resources—its faculty, parents, and students.

Ultimately, each level of the system must assume its appropriate share of responsibility:

- The *state* and the *district central office* must each be accountable for the policies they adopt: ensuring that their decisionmaking processes rely on the best available knowledge and the interests of children, and evaluating the subsequent effects of the policies—on teaching and learning, on student access to opportunities, on time for teaching, on the extent of paperwork created, etc. They must assume responsibility for equity in the distribution of school resources, and for creating

processes that keep them responsive to the needs and concerns of parents, students, and school-level staff.

- The *school* must be accountable for adopting policies that reflect professional knowledge and the interests of children, for equity in the distribution of learning opportunities to children, for establishing means by which staff can continually learn and refine their skills, for creating processes for communication as well as for problem identification and problem-solving, and for creating methods for involving and responding to parents, students, and staff concerns and ideas.

- *Teachers and other staff* must be accountable for identifying and meeting the needs of individual students responsibly and knowledgeably based on professional standards of practice, for continually evaluating how well their practices are accomplishing this goal, for seeking new knowledge and information, and continually revising their strategies to better meet the needs of students.

An accountability system helps everyone to do his or her job more responsibly by providing information about school practices and outcomes along with occasions for learning, consultation, and problem-solving.

Creating Processes that Enhance School Accountability

Professionals are obliged to do whatever is best for their clients based on the best available knowledge in their fields. Thus, at the school level, where good education is defined as meeting the needs of individual students, arguably the most important form of accountability is professional accountability.

Professional accountability seeks to support practices that are learner-centered and knowledge-based. This form of accountability focuses on strategies that are intended to ensure that:

1) All individuals permitted to practice are competent to do so responsibly;

2) where knowledge about appropriate practice exists, it is used in making decisions;

3) where certainty about practice (or about the best solution to a particular student need or problem) does not exist, practitioners—individually and collectively, through inquiry and consultation—continuallyseek to discover the most responsible course of action.

In contrast to bureaucratic accountability, which can only ensure uniform procedures, professional accountability seeks to ensure responsible decisionmaking. It exchanges regulations that prescribe what is to be taught, when, and how for greater investments in teacher competence. These investments are made through rigorous and meaningful education, certification, selection, and evaluation, along with opportunities for ongoing learning, shared inquiry, and consultation (Darling-Hammond, 1990c). A professional accountability strategy is needed in education because the complexities of teaching and learning demand that educators be trusted—and be trustworthy—to make difficult decisions about the special, infinitely variable needs of individual students.

To create professional accountability, a system must pay particular attention to the following:

- *personnel policies* governing the hiring, assignment, and evaluation of teachers and other staff;
- opportunities for *professional development*, *shared inquiry*, and *consultation about problems of practice* (means for fostering the use of professional knowledge beyond that represented in the experiences of individual teachers);
- *assessment vehicles* for continually evaluating classroom and school practices, as well as student progress;
- the creation of *incentives* to sustain teacher (and other staff) learning, self-evaluation, and a willingness to assume teaching and leadership challenges within the classroom and the school.

Collective problem-solving is a second major aspect of school accountability. It is the effort to establish an inquiry ethic and a commitment to identify, analyze, and address issues or problems that might impede student and school progress. Such an ethos must be supported by methods for continually evaluating what is going on, asking not just what is occurring but also why it is happening and whether existing practices are accomplishing what the school community wants to accomplish. These questions should not be raised only once or twice a year, when the students are tested or the annual needs assessment form is filled out. Rather, they should be raised in every faculty and team meeting, on every occasion when faculty and students are striving to meet their goals, at every juncture when any kind of stock-taking occurs.

In a few schools, this kind of collective questioning and reflection is frequent. More commonly, teacher isolation has worked against collective accountability, while centralized planning, decisionmaking, and evaluation have often removed occasions as well as incentives for this kind of activity at the school level.

Yet, if schools are to become more responsive and open to change, they must find ways—as other professional organizations have done—to make evaluation and assessment part of their everyday activities. Just as hospitals have standing committees of staff that meet regularly to discuss the effectiveness of various aspects of the hospital's functioning (surgery, pathology, epidemiology, etc.), so schools must have regularized occasions for examining their practices. And just as lawyers, doctors, psychologists, social workers, and other professionals use case conferences and other forms of consultation as opportunities to share knowledge and solve problems on behalf of their clients, teachers also must have opportunities to profit from their colleagues' knowledge and perspectives on behalf of their students.

The first type of evaluation, teacher involvement in peer review of practice, can take several different forms. Standing committees, such as those used in hospitals, can meet regularly to review practices in different areas—parent-school relations, student grading and promotion policies, academic progress, the quality of particular programs or curricula, the organization of instruction, and so on. Ad hoc committees can be formed to work on special issues. Faculty meetings might be used to investigate curricular strategies and other matters within and across departments or grade levels. These regularized evaluation activities provide opportunities for using assessment data and other feedback to inform decisionmaking.

The creation of consultation structures provides a second mechanism for accountability. Structures can include opportunities for staff to review an individual student's progress or for collective assessments of how to help students with difficulties. They can also include joint planning activities of faculty teams to foster collegial relationships among teachers that support consultation about curriculum concerns, teaching methods, and problems of practice.

Central Park East, a New York City elementary school, uses both peer review of practice and collective consultation mechanisms to ensure that school and student

concerns are continually aired, and that collaborative problem-solving occurs. Among many other activities, the staff have two different weekly meetings for this purpose. One meeting is used either for a student review or a curriculum review. In the first instance, a teacher discusses the status of an individual student, and faculty members add their observations about the child, about teaching strategies or interpersonal approaches that may be useful, and about research or experience that pertains to some of the issues being raised. The process enriches everyone's understanding about both the particular child under discussion and about teaching in general. Curriculum review meetings use a similar strategy to examine a particular curricular issue of concern in the school.

The other weekly meeting deals with schoolwide concerns, such as multicultural education, parent-school relations, or school assessment. Here, as is the case with the student or curriculum review, faculty members' observations, information collected in the classroom, school, or home setting, and pieces of assessment data (at the student, classroom, or school level) may ground certain aspects of the discussion. These and other processes help members of the school community assume responsibility for the welfare of students and for the life of the school. They promote genuine accountability by creating an engine for continual reflection, evaluation, and change.

Creating Indices That Enhance School Accountability

To support a school's evaluative efforts, a school level indicator system must provide useful information about the school context, that is, what the school is doing, as well as how well students are achieving.

Measures of Student Learning. As "outcome" measures, most commercially published standardized tests are not sensitive to differences in school curricular offerings, and so do not really assess the opportunities for learning that students are afforded in various classrooms and schools (Madaus et al., 1979). To elicit a better measure of a school's effectiveness, there are several options. Most conventional are the curriculum-based tests constructed by some states and school districts, which reflect more precisely curriculum goals and variations in school teaching (Madaus et al., 1979).

In addition, a number of less traditional options are also being developed, including portfolios, exhibitions, demonstrations, and other performance-based

assessments. In contrast to multiple-choice standardized tests, these assessment strategies present ill-structured problems that require students to think analytically and demonstrate their proficiency as they would in a real-life performance situation. (Archbald & Newmann, 1988).

Much like the kinds of assessments that prevail in most other countries around the world (where multiple-choice testing has not gained much popularity), these approaches include essay examinations, research projects, scientific experiments, oral exhibitions, and performances in areas like debating and the arts. They also include portfolios of students' best work in various subject areas, and group projects that require analysis, investigation, experimentation, cooperation, and written, oral, or graphic presentation of findings. Often the assessment occasion requires students to respond to questions from classmates or from external examiners, thus helping them learn to think through and defend their views (Coalition of Essential Schools, 1990; Archbald & Newmann, 1988).

In addition to helping teachers and the students themselves evaluate what the students can *really* do, these approaches serve as expressive tools for students and are highly motivating. Sizer points out that they are as much inspiration as measurement: "Giving kids a really good target is the best way to teach them... And if the goal is cast in an interesting way, you greatly increase the chances of their achieving it" (Coalition of Essential Schools, 1990, p. 1).

A number of schools, such as those in the Coalition of Essential Schools, are engaged in creating authentic assessments of student learning. In addition, states like Vermont, California, Connecticut, and New York are beginning to experiment with new forms of assessment. Vermont is developing student portfolios as the basis of its state assessment system. Connecticut and New York have begun to develop performance-based assessments that require students to perform a science experiment or solve a real world problem using mathematical and scientific concepts rather than to complete a multiple-choice test. Maryland has developed a writing assessment that engages students in complex writing tasks, sometimes requiring several days of work, including revisions, as part of the examination process. Districts such as Shoreham-Wading River, New York; Pittsburgh, Pennsylvania; and Albuquerque, New Mexico, are also creating authentic assessments to take the place of standardized testing. An Urban

School District Consortium recently launched by the American Federation of Teachers has a large and growing list of members collaborating in the development of performance assessments. A special federal task force has been convened to examine alternatives to standardized testing for Chapter 1 compensatory education programs.

Of course, performance tests must still address the problems that have confronted previous generations of standardized tests. First, although studies have shown that raters can achieve a high degree of agreement about essays and portfolios, the problems of cultural bias that also plague existing standardized tests remain unresolved. In fact, there are indications that minority students may do better on some kinds of performance tests and more poorly on others, depending on the ways in which tasks have been selected and on the scoring criteria and processes. Second, the time and training needed to score performance tests will surely seem less efficient when compared to the 10,000 mechanically scored tests per hour that the Educational Testing Service now achieves: states and school districts that are strapped for funds may be hard pressed to choose performance tests until their benefits are widely viewed as justifying the costs. Third, although these tests are being mandated to enrich teaching and learning, the very pressure of any "high stakes" program could work against these goals. Depending on the nature of assessment tasks, some evidence suggests that instruction can be geared quite narrowly to the format or content of performance assessments, just like other tests (Ascher, 1990). To fulfill their promise, performance assessments will need to be constructed to draw on a wide range of analytic skills and higher-order conceptual abilities, as well as on developed performance capacities, so that teaching *for* them can only be accomplished by maintaining broad and enriched goals for student learning.

Regardless of which testing instruments schools may currently need to use to satisfy existing state or federal requirements, local school staff should examine how they can be more genuinely accountable for valued student learning by creating means for examining what students are actually able to do: demonstrations of mastery that respond to students' desires to achieve and to show their competence, and to the community's long-range goals for its students.

Measures of School Performance. To inform school decisionmaking and improve practice, assessments of what students are learning must be accompanied by

assessments of what the school is doing. In the previous section, some school context variables important to school effectiveness were discussed. They include features that control students' access to knowledge, the school's press for achievement, and professional teaching conditions. With creativity, a set of indicators can be devised for each one. For example, the press for achievement might be elicited by measures concerning the proportion of students enrolled in rigorous programs, the school's recognition of academic accomplishments, the extent to which time for teaching and learning is safeguarded from interruptions or fragmentation, and the quality and type of homework assigned in various classes. Using a number of such indicators for each of these three areas is a good way to begin to evaluate the quality of education provided by an individual school.

School indicators should look at the school environment holistically, describing the quality of children's school experiences. Indicators can consist of systematic observations by faculty of student behavior or of classroom and school practices; information derived from surveys, interviews, or focused conversations with students, parents, or staff; counts of particular events or aggregations of recorded data; and structured evaluations of school or student performance.

Indicators of the school's performance should always start with questions of how students are doing: their attendance and sense of connection to school, resourcefulness, self-esteem, sense of responsibility, civic and social growth, and other desired attributes along with academic achievement. However, these indicators should be expanded to reflect factors that structure students' experiences of schooling:

- Who teaches what to whom? (What is the distribution of teaching expertise and curricular opportunity?)
- How is time allocated across activities and subject areas?
- What opportunities are there for each child to have success and build on strengths?
- What kinds of extra- or co-curricular opportunities are available? How many children are able to participate?
- How are teachers and students grouped for instruction? What are the probabilities that each child will have an adult advocate who knows him or her intimately and can be responsible for his or her welfare?

- How and how often are parents involved in school activities and decisions? What decisions are students involved in making? What about teachers and other staff?

- How is class time used? How much time is spent on lecture, seatwork, student projects, laboratory work, writing, problem-solving? What kinds of tasks are students asked to perform?

- What kinds of intellectual material do students encounter? Does this vary by classroom, group, or track?

- What do students think about their school experiences? What aspects of their experience do they find motivating? What aspects are discouraging?

- What do parents think about their children's school experiences?

- What aspects of students' home experiences are important contributors to the students' schooling experience as well as to the school's efforts to make connections with parents? (e.g., family structure and circumstances, distance from school, student employment outside of school, and so on).

With questions like these driving the development of indicators, school faculty can use their findings to make well-grounded decisions, to identify areas in which they want to work on change, and to establish benchmarks for tracking their successes.

Going within and beyond the School as a Unit

Although school-based analyses can provide rich data, for some educational issues the school is either too small or too large a unit of analysis. For example, schools with tracks and other educational groupings offer students greatly varied opportunities to learn. Thus, it is important to know not only whether a school offers calculus, for example, but which students can take this course—and what the mathematics options are for those students who cannot enroll in calculus. Similarly, is the training of teachers equally good in general track classes as in classes serving the academic track? In the same vein, while a school may contain computer equipment, computers may be differentially available to students in a general and academic track. Moreover, the ways in which computers are used—for drill, or for programming and problem-solving—may differ depending on the track.

At the same time, to answer questions of equity, such as desegregation or school finance, one must go beyond school level data. For example, there is good

evidence that state education formulas set urban systems at a disadvantage in some states and poor rural districts at a disadvantage in others. Within urban districts, resource allocations also tend to discriminate against the poorest (and most nonwhite) schools (Ascher, 1989; Fruchter, 1989). Thus, for parents to be sure that their child is receiving equitable resources, the resources available at their school must be compared to those at other schools, in the district and beyond.

Indicators as Incentives

Research consistently suggests that people will do more in the areas in which they are evaluated. This is why assessments are so powerful. Indicators not only measure reality, but they change it. When indicators come with stakes attached—school accreditation, financial incentives, student promotion or graduation, and so on—changes in behavior can be predicted with great certainty (Haney & Madaus, 1986). Depending on the circumstances, the type of measure, and the stakes attached, the indicators may cease to measure what they were originally intended to measure (Darling-Hammond, 1988).

As noted, the effects of high stakes indicators can sometimes be counterproductive. For example, when teachers and principals are given rewards according to student outcomes averaged at the school building level, two immediate incentives are created: for schools to try to keep out, or push out, low-scoring students; and for talented teachers and administrators to transfer out of schools serving educationally disadvantaged students. Obviously, neither outcome serves the goals of the policy. Instead, schools become less accountable rather than more so.

On the other hand, well-conceived indicators that are appropriately used can serve as positive incentives for school improvement. If indicators of student learning encourage students and teachers to focus on important skills and abilities, if they help identify needs or problems so that they can be better addressed, then they will support responsive schooling. Similarly, if indicators of school context or performance help faculties and school communities to monitor the quality and the equality of opportunities made available to students, if they provide rich information for school-based problem-solving, then they will support responsible decisionmaking and, in the long run, more accountable education.

REFERENCES

Archbald, D., & Newmann, F.M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school.* Reston, VA: National Association of Secondary School Principals.

Ascher, C. (1990). *Can performance-based assessments improve urban schooling?* Digest No. 66. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University.

Ascher, C. (1989). *Urban school finance: The quest for equal opportunity.* Digest No. 55. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. (ED 311 147)

Boyer, E. (1983). *High school.* New York: Harper and Row.

Carnegie Council on Adolescent Development. (1989). *Turning points: Preparing youth for the 21st century.* New York: Carnegie Corporation of New York. (ED 312 322)

Coalition of Essential Schools. (1990, March). Performances and exhibitions: The demonstration of mastery. *Horace, 6* (3).

Darling-Hammond, L. (1990a). *Indicators and policy.* Paper prepared for the Organization of Economic Cooperation and Development. Paris: OECD.

Darling-Hammond, L. (1990b). Teacher quality and equality. In J. Goodlad & P. Keating (Eds.), *Access to knowledge: An agenda for our nation's schools.* NY: College Entrance Examination Board.

Darling-Hammond, L. (1990c). Teacher professionalism: Why and how? In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now.* Bristol, PA: Falmer Press, p. 25-50.

Darling-Hammond, L. (1989, Fall). Accountability for professional practice. *Teachers College Record.*

Darling-Hammond, L. (1988). *Assessment and incentives: The medium is the message.* The AAHE Assessment Forum. Third National Conference on Assessment in Higher Education, Chicago, June 8-11.

Darling-Hammond, L. & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85* (3), 315-336.

Fruchter, N. (1989, Summer). Rethinking school reform. *Social Policy, 20* (2), 16-25.

Goodlad, J. (1984). *A place called school: Prospects for the future.* New York: McGraw Hill. (ED 236 137)

Haney, W., & Madaus, G. (1986). Effects of standardized testing and the future of the national assessment of educational progress. Working paper prepared for the NAEP study group. Chestnut Hill, MA: Center for the Study of Testing, Evaluation and Educational Policy.

Holmes, C.T., & Matthews, K.M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*, 225-236.

Koretz, D. (1989). A framework for evaluating and validating indicators of mathematics and science education. Santa Monica, CA: RAND Corporation.

Koretz, D. (1988, Summer). Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12* (2), 8-15, 46-52.

Lieberman, A., & Miller, L. (1990, June). Restructuring schools: What matters and what works. *Phi Delta Kappan, 71* (10), 759-64.

Madaus, G.F., Kellaghan, T., Rakow, E.A., & King, D.J. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review, 49* (2), 207-229.

McDonnell, L.M. (1989). *Restructuring American schools: The promise and the pitfalls.* New York: Institute on Education and the Economy, Teachers College, Columbia University. (ED 314 547)

McDonnell, L., & Oakes, J. (1988). *Creating responsible and responsive accountability systems.* Report of the OERI Study Group on State Accountability Reporting. Santa Monica, CA: RAND Corporation.

National Assessment of Educational Progress (NAEP). (1981). *Reading, thinking, and writing: Results from the 1979-80 National Assessment of Reading and Literature.* Denver: Author.

Oakes, J. (1989, Summer). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis, 11* (2), 181-199.

Oakes, J. (1986, October). *Educational indicators: A guide for policymakers.* New Brunswick: Rutgers University, Center for Policy Research in Education. (ED 315 920)

Oakes, J., Ormseth, R., Bell, R., & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class, and ability grouping on access to science and mathematics education.* Santa Monica, CA: RAND Corporation.

OERI State Accountability Study Group. (1988). *Creating responsible and responsive accountability systems. Report of the OERI State Accountability Study Group.* Washington, DC: U.S. Department of Education, Office of Educational Research and Information. (ED 299 706)

U.S. Department of Education. (1987). *White paper on accountability: Tying assessment to action*. Washington, DC: Author.

Wise, A.E., & Gendler, T. (1989). Rich schools, poor schools: The persistence of unequal education. *The College Board Review, 151,* 12-27.

LINDA DARLING-HAMMOND is the Co-Director of the National Center for Restructuring Education, Schools and Teaching, Teachers College, Columbia University. She was previously Senior Social Scientist and Director of the Education and Human Resources Program at the Rand Corporation. She is co-editor of *The Second Handbook of Teacher Evaluation*, and author of more than 80 journal articles, book chapters, and monographs on educational policy issues. Dr. Darling-Hammond received her doctorate in urban education from Temple University.

CAROL ASCHER is Senior Research Associate at the ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. She is the author or editor of four non-fiction books and the author of a novel, *The Flood*. She has also written more than 50 monographs covering a wide range of urban education issues. Dr. Ascher received her doctorate in anthropology from Columbia University.

ERIC Clearinghouse on Urban Education
Institute for Urban and Minority Education
Box 40, Teachers College, Columbia University
New York, New York 10027

National Center For Restructuring Education, Schools and Teaching
Box 110, Teachers College, Columbia University
New York, New York 10027